# VinBigdata Chest X-ray Abnormalities Classification

Trung Phan Le Chi [1,2*], Man Nguyen Tran [1,2*], Hao Nguyen Phuc [1,2*], Duy Do Phu [1,2*], Luong Nguyen Ngoc [1,2*] and Thien Nguyen Tat Bao [1,2,3*]

[1*]The University of Information Technology Vietnam National University - Ho Chi Minh City (VNU-HCM).
[2]Faculty of Information Science and Engineering.
[3]Instructors.

*Corresponding author(s). E-mail(s): 21522725 @uit.edu.vn; 21522325 @uit.edu.vn; 21522047 @uit.edu.vn; 21520205@uit.edu.vn; 21522311 @uit.edu.vn; thienntb@uit.edu.vn;

**Abstract**

This paper presents an innovative approach to medical image diagnosis leveraging the **VinBigData Chest X-ray Abnormalities Detection** dataset. With the increasing volume of medical imaging data, the need for accurate and efficient diagnostic tools is paramount. Our study employs advanced deep learning techniques, including deep neural networks, to detect abnormalities in chest X-ray images. We explore various architectures and methodologies to enhance diagnostic performance while considering computational efficiency. Through rigorous experimentation and evaluation, our proposed model achieves state-of-the-art performance in detecting a range of abnormalities, including pneumonia, lung nodules, and other thoracic pathologies. Besides that, our research explored the effection of preprocessing to each diseases. The findings of this research hold significant promise for improving healthcare outcomes by providing reliable, scalable, and accessible diagnostic solutions for medical professionals.

**Keywords:** VinBigData Chest X-ray, Swin, Unet, VGG-19

# 1 Introduction

The rapid and accurate diagnosis of chest abnormalities is crucial, as some paroxysmal diseases may cause death within a very short time. Medical imaging, as a widespread screening method, is of great significance for thoracic anomaly detection, particularly for detecting heart and lung diseases and skeletal abnormalities. Chest X-ray, as one of the most common types of medical imaging, is usually the first radiological examination for clinical diagnosis. Indeed, chest X-rays have remarkable advantages, such as having a low cost, fast imaging speed, small amounts of radiation, and reasonable sensitivity to various pathologies, allowing them to play a vital role in common screening and emergency treatments, such as heart failure, pneumonia, pulmonary edema, pulmonary nodules, pleurisy, and chest foreign bodies.

Currently, chest X-rays are mainly read manually by professional radiologists, who are primarily limited by their experience, efficiency, and the complexity of the image itself. The analysis and recognition of chest X-rays are formidable challenges, because their anatomical structures and pathological features are extremely complicated. It is difficult to identify abnormalities in certain parts, especially multi-scale abnormalities located in indistinguishable regions with different appearances. Therefore, observing and analyzing chest X-rays are time-consuming and labor-intensive tasks for radiologists. With the development of artificial intelligence and big data, data-driven computer vision technology has played an increasingly significant role in assisting radiologists in the detection and diagnosis of illnesses. The intelligent computing and analysis of medical images cannot only effectively relieve the working pressure of radiologists, but also increase the diagnosis speed and improve sensitivity to subtle abnormalities. Considering the interpretation complexity and clinical value of chest X-ray images, many scholars have been inspired to research automated algorithms for identifying heart and lung diseases.

Recently, deep learning technology has demonstrated the advantages of the automatic and efficient analysis of big data, which is widely used in various fields, including medical image analysis. Numerous variants of convolutional neural networks (CNNs) have been developed. These variants achieved remarkable performances in numerous medical analysis tasks and were comparable to professional radiologists. Specifically, many advanced CNN algorithms have demonstrated promising performances in the interpretation of chest X-rays. Based on the type of visual label, a chest X-ray analysis can be roughly divided into image-level prediction, image segmentation, and instance-level detection. First, image-level prediction is the process of predicting a category label or set of consecutive values for the entire image. The classification labels can include some common abnormalities, such as pneumonia and emphysema, and the regression value may indicate the severity score of a particular anomaly. In terms of image-level prediction, a representative study of chest X-rays was performed on the ChestX14 dataset. Some studies used classification networks such as ResNet and DenseNet to predict anomaly labels of the Chest X-ray14 dataset, and then showed abnormal regions through visualization methods, such as class activation mapping (CAM) or gradient-weighted class activation mapping (Grad-CAM). These strategies can improve the interpretability of the network to some extent, but they cannot be

used for an accurate quantitative analysis of the anomaly localization accuracy. Second, chest X-ray segmentation refers to the labeling of each pixel in the image. In the chest X-ray imaging domain, segmentation targets can include organs, anomalies, the lungs, skeleton, and external objects. These tasks usually segment only one object of interest, and all remaining pixels are marked as the background, because accurately annotating multiple outlines of chest abnormalities in X-ray images is difficult and time-consuming. Finally, chest X-ray instance-level prediction refers to the labeling of one or multiple specific regions in a chest X-ray image. These regions are typically rectangular boxes. This prediction pays more attention to the detection of anomalous locations and provides rapid and practical assistance for clinical diagnosis. In this study, we aim to improve deep networks and achieve more accurate instance-level predictions under supervised learning.

Therefore, to address the above problems, this study is devoted to the multi-label and instance-level anomaly classification of chest X-rays. Using a dataset of 15,000 scans along with the corresponding label annotated by experienced radiologists, we have applied different networks for training and comparing. With this method, an accurate identification can be generated for a certain chest X-ray image, thus relieving the stress of busy doctors while also providing patients with a more accurate diagnosis.

.

## 2 Related works

Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T., & Saalbach, A. (2019): compares deep learning architectures (DenseNet, ResNet, etc.) and training strategies for multi-label chest X-ray classification, analyzing the impact of data and image quality. Li, Y., Zhang, Z., Dai, C., Dong, Q., & Badrigilan, S. (2020): analyzes prior studies on deep learning for chest X-ray pneumonia detection, evaluating accuracy across various datasets and highlighting factors influencing model performance. In 2021, Huang, H., Long, Y., & Wei, Y. : compares the performance of the proposed method with other deep learning algorithms for chest X-ray abnormality detection, using the VinBigData dataset from Kaggle. The method is also evaluated on different chest X-ray datasets and demonstrates good generalization ability. Taslimi, S., Taslimi, S., Fathi, N., Salehi, M., & Rohban, M. H. (2022): proposes a novel transformer-based model (Swin Transformer) for multi-label chest X-ray classification, achieving superior results on ChestX-ray14 and advocating for a standardized evaluation approach. Zouch, W., Sagga, D., Echtioui, A., Khemakhem, R., Ghorbel, M., Mhiri, C., & Hamida, A. B. (2022): reviews deep learning for COVID-19 detection in CT scans and chest X-rays, exploring architectures, data augmentation, and limitations of existing methods. Sharma, S., & Guleria, K. (2024): provides a comprehensive review of deep learning approaches for pneumonia detection in chest X-ray images. It discusses various CNN architectures, data augmentation techniques, transfer learning, and ensemble methods used for this task.

# 3 Dataset

## 3.1 Source and Crawling Method

The dataset was provided by two hospitals in Vietnam: the Hospital 108 and the Hanoi Medical University Hospital for the Kaggle competition - VinBigData Chest X-ray Abnormalities Detection.

It can be found online at : VinBigData Chest X-ray Abnormalities Detection

## 3.2 Annotation Process

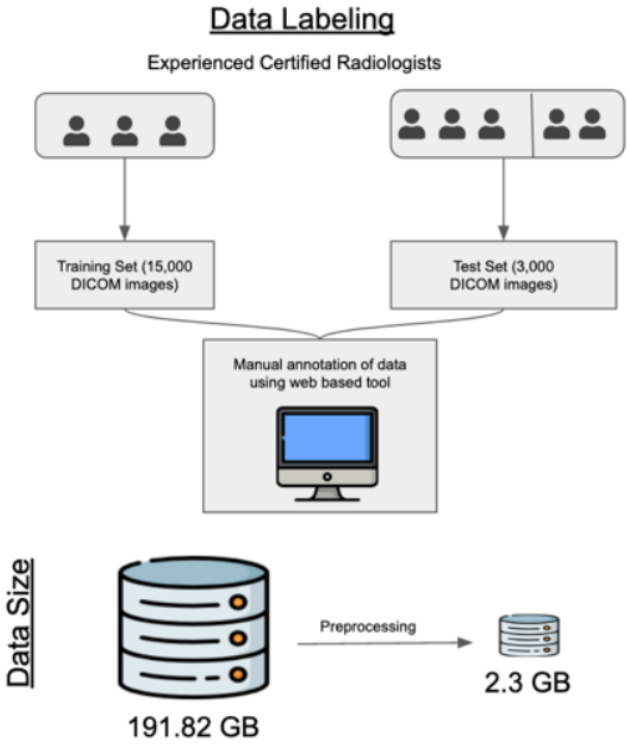


**Fig. 1**: Overview of data labelling process

A total of 17 different radiologists were involved in manually annotating this dataset. The entire dataset was further divided into a training set of 15,000 images and a test set of 3,000 images. Each scan in the training set was independently labeled by 3 radiologists while each scan in the test set was labeled by consensus of 5 radiologists.

All images were labeled for the presence of 14 critical radiographic findings, which are aortic enlargement (0), Atelectasis (1), Calcification (2), Cardiomegaly (3), Consolidation (4), ILD (5), Infiltration (6), Lung Opacity (7), Nodule/Mass (8), Other

lesion (9), Pleural effusion (10), Pleural thickening (11), Pneumothorax (12), and Pulmonary fibrosis (13). Each number in the parentheses represents the corresponding label number. Note that the "No finding" observation (14) captured the absence of all 14 findings above.

| **Column Name** | **Column Description** |
|---|---|
| Image_id | Unique image identifier |
| class_name | Name of abnormality present in a specific image |
| class_id | Unique numerical ID (from 0 - 14) assigned to each class in the dataset |
| rad_id | ID of the radiologist (from 1 - 17) that annotated that image |
| x_min, y_min, x_max, y_max | The last 4 columns contain the bounding box coordinate information for the exact location of the abnormality in an image where (x_min, y_min) tuple represents top left corner and (x_max, y_max) tuple represents bottom right corner of the bounding box. For the "No Finding" class, these columns have NaNs as placeholders. |

**Table 1**: The description of each column

In train datasets, one row for each object includes a class and a bounding box. Note that some images have multiple bounding boxes. For each test image, we will be predicting bounding boxes and classes for all findings. If no findings are predicted, a prediction of "14 1 0 0 1 1" should be created (14 is the class ID for no finding, and this provides a one-pixel bounding box with a confidence of 1.0). Each row contains information about a single annotation made by a single radiologist on a single image. Furthermore, each image appears at least 3 times in the image_id column because each image was independently labeled by 3 different radiologists. The description of each column is given in table [1]:

### 3.3 Dataset Statistics

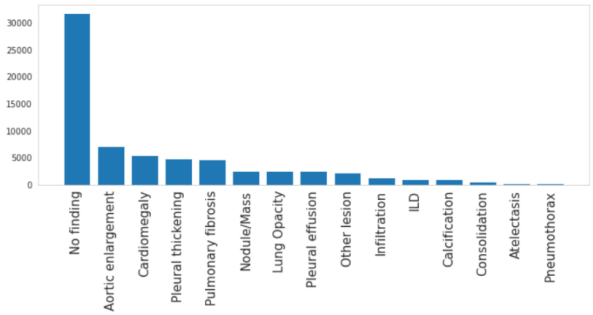


**Fig. 2**: The distribution of the labels of the dataset

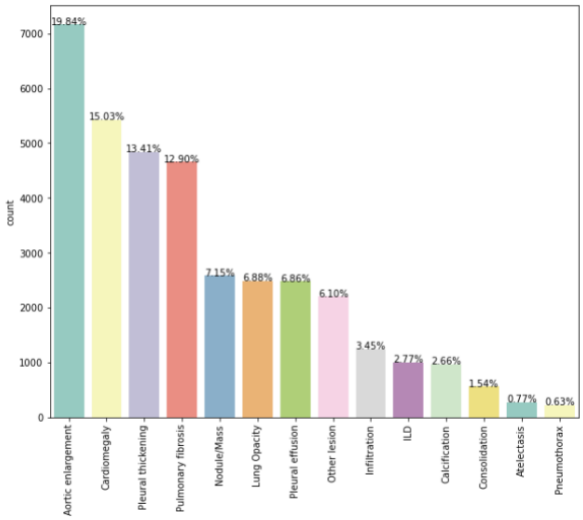


**Fig. 3**: The distribution of the labels of the dataset without 'No finding' label

Because this dataset belongs to a challenge on Kaggle so the testing set will not have labels so we only use the training set in this paper. The training set that we use comprises 15,000 postero-anterior (PA) CXR scans in DICOM format, which were de-identified to protect patient privacy. The distribution of the labels of the dataset is shown in Figure 2. We could immediately see that the dataset is imbalanced with the 'No finding' label that contains over 30000 values. This label is used for the detection task which includes bounding boxes and we only perform the multilabel classification task so this label will be removed when training models and the new distribution of the labels in the dataset can be seen in Figure 3. Aortic enlargement is the most

popular disease beside Cardiomegaly, Pleural thickening and Pulmonary fibrosis. On the other hand, Consolidation, Atelectasis and Pneumothorax are rarely to be found. Some examples of the dataset are presented in Figure 4. We could see that the red rectangle are the bounding boxes and each of them will come up together with a label of illness. With a more comprehensive observation, the locations of the bounding boxes are different with the corresponding diseases. This raised up a problem of preprocessing with will be presented in the sections below. The dataset is then divided into training, validation and testing sets with the ratio of 8:1:1.

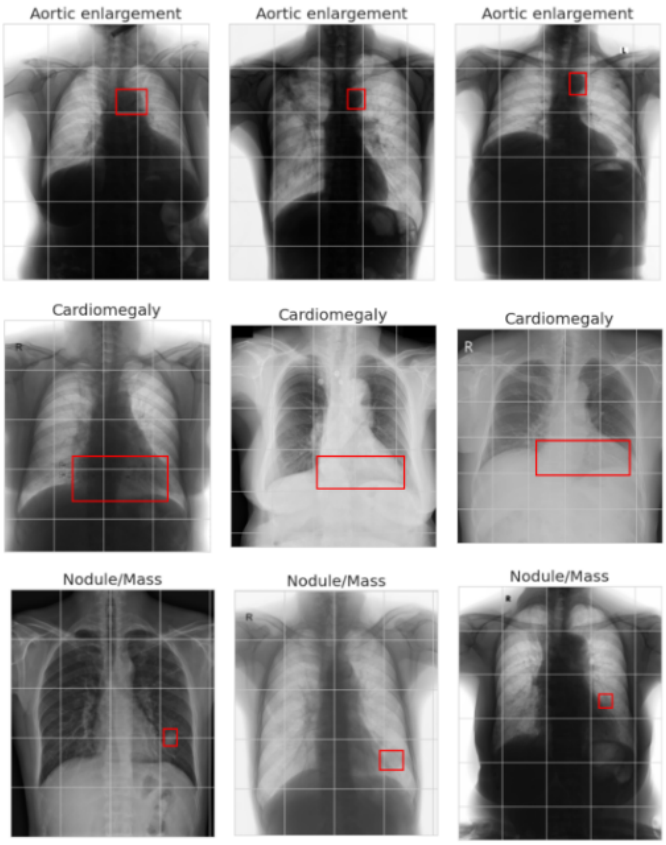


**Fig. 4**: Some examples of the dataset

# 4 Preprocessing methods

In biomedical image processing, highlighting 2D "edges" is crucial for focusing on specific features. Image gradients, which capture the change in pixel intensity, are particularly effective for this task, as they pinpoint areas where color values transition sharply. Our work investigates the effectiveness of various image preprocessing

techniques on a 15-class classification task. By comparing the model's performance on each label with different preprocessing methods, our goal is to be able to identify the most suitable preprocessing approach for each class. We decided to do the experiment on different methods including Exposure - Histogram, Laplace - Gaussian, Sobel - Feldman and Thresholding.

## 4.1 Increasing exposure, adjusting histogram

Preprocessing medical images often involves histogram equalization, a technique found in libraries like scikit-image (exposure.equalize_hist). This method acts like boosting image exposure. By spreading out the pixel intensities, it brightens dark areas and clarifies light ones, enhancing overall contrast. This can improve the visibility of subtle details crucial for medical image analysis and potentially lead to better model performance.

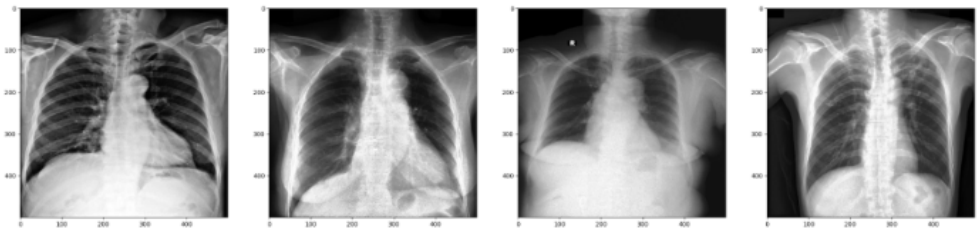

**Fig. 5**: Examples of Increasing exposure, adjusting histogram

## 4.2 The Laplace filter with Gaussian second derivatives:

The Laplacian filter, also known as the Laplace-Gaussian filter, utilizes second derivatives calculated with a Gaussian kernel. This approach targets pixels with significant intensity changes. Furthermore, by incorporating Gaussian smoothing beforehand, the filter effectively removes noise while preserving sharp edges.

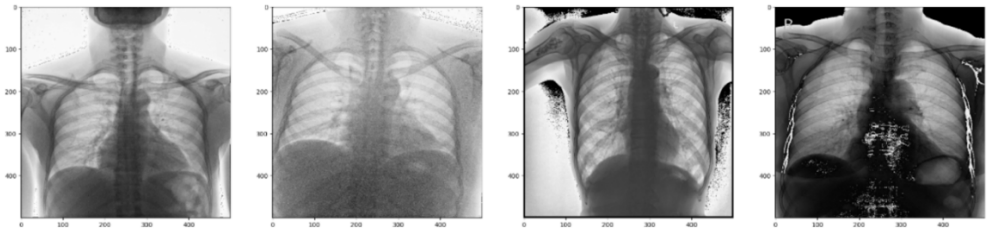

**Fig. 6**: Examples of The Laplace filter with Gaussian second derivatives

## 4.3 The Sobel-Feldman operator (the Sobel filter)

The Sobel-Feldman operator, also known as the Sobel filter, is a technique well-suited for detecting edges in 2D X-ray images. It works by finding regions of high spatial frequency, essentially the edges themselves. This method utilizes two small (3x3) kernel matrices, one for the horizontal axis and another for the vertical. These kernels are applied to the X-ray image through a convolution operation. Finally, the gradients obtained from both axes are combined using the Pythagorean theorem to calculate a single gradient magnitude, representing the strength of the edge at that point.

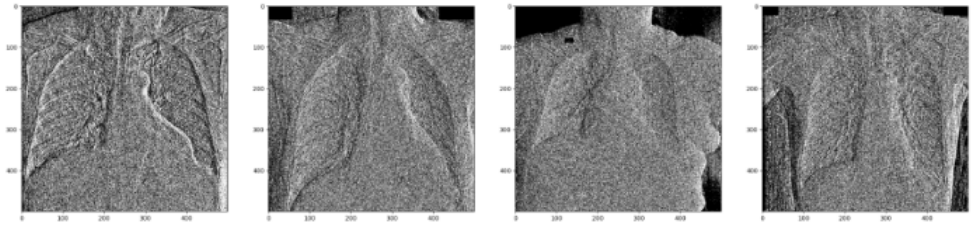

**Fig. 7**: Examples of the Sobel filter

## 4.4 Apply masks to X-rays with thresholding (Thresholding method)

In X-ray image analysis, isolating specific features often proves beneficial. This can be achieved by applying masks, which are binary arrays that define which pixels to keep or discard. We explored two ways to separate an X-ray image into light and dark areas (foreground and background) using a threshold value of 150:

### 4.4.1 Keeping Details (Informative)

- Pixels brighter than 150 (likely interesting areas) keep their original brightness in the new image.
- Darker pixels (background) are turned completely black
- This approach helps maintain details in brighter areas for further analysis.

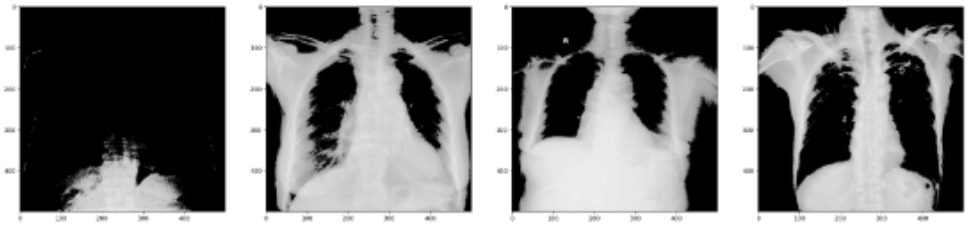

**Fig. 8**: Examples of Keeping Details

### 4.4.2 Simpler Separation (Binary)

- Bright pixels ($> 150$) become white (1) in the new image.
- Darker pixels are still black (0).

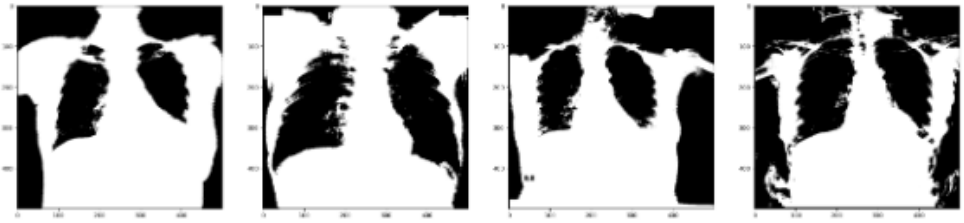

**Fig. 9**: Examples of Simpler Separation

# 5 Method

## 5.1 Swin Transformer

The key feature of the SwinTrans or Swin Transformer is its use of non-overlapping windows that shift between layers. This method allows the model to efficiently capture both local and global visual features by processing image patches within windows and then shifting these windows to capture cross-window connections. This hierarchical design makes the Swin Transformer scalable and suitable for a wide range of vision applications, from image classification to object detection and segmentation

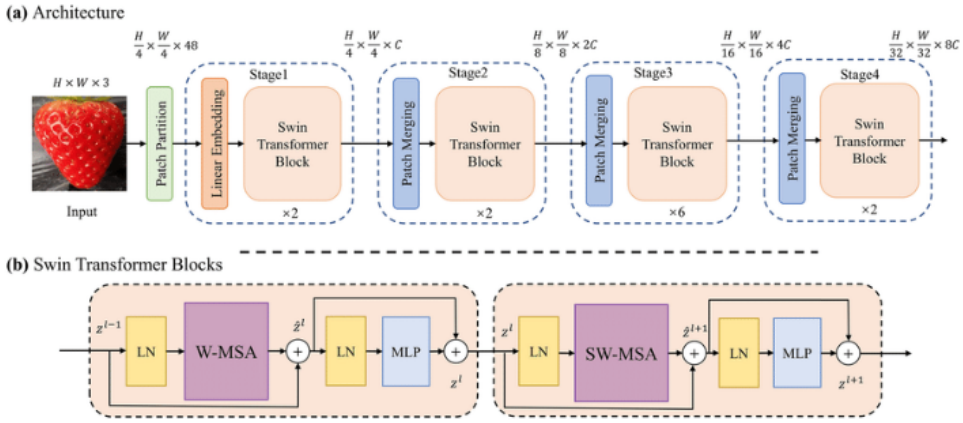


**Fig. 10**: Swin Transformer architecture

## 5.2 CoAtNet

CoAtNet is a neural network architecture that combines the strengths of Convolutional Neural Networks (CNNs) and Transformers. It is designed to leverage the local feature extraction capabilities of CNNs and the global context modeling power of Transformers to achieve superior performance in computer vision tasks.

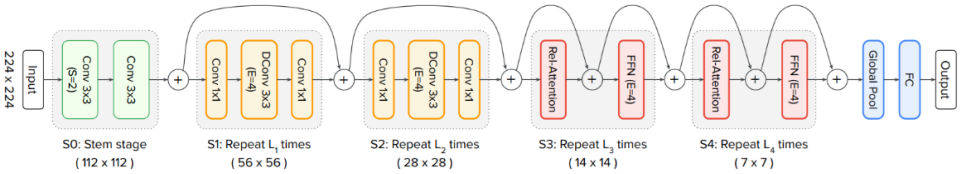


**Fig. 11**: CoAtNet architecture

## 5.3 Unet Transformer

UNETR, or UNet Transformer, is a Transformer-based architecture for medical image segmentation that utilizes a pure transformer as the encoder to learn sequence representations of the input volume – effectively capturing the global multi-scale information. The transformer encoder is directly connected to a decoder via skip connections at different resolutions like a U-Net to compute the final semantic segmentation output.

Although UNet Transformer was originally designed for image segmentation, with appropriate adjustments it can also be used for image classification. In this paper, we will change the output layer and loss function to suit the problem and evaluate its medical image classification ability.

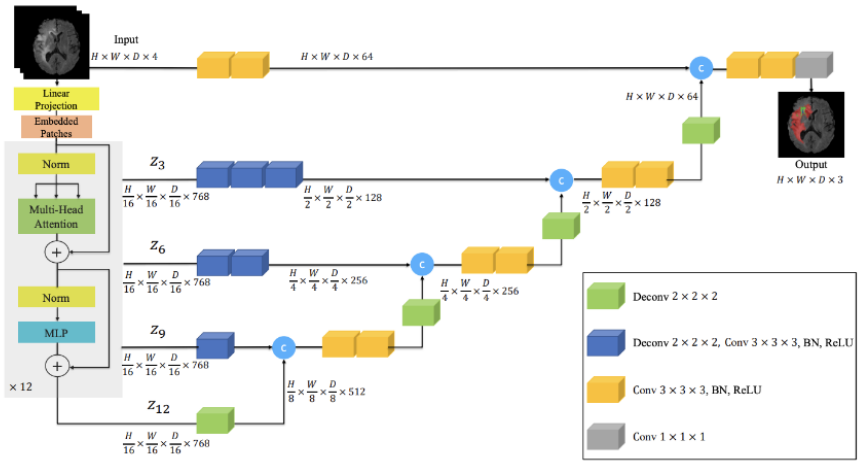


**Fig. 12**: Unet Transformer architecture

## 5.4 VGG-19

VGG-19 is a convolutional neural network (CNN) model developed by the Visual Geometry Group (VGG) at the University of Oxford. It's renowned for its depth and simplicity, and it played a pivotal role in the advancement of deep learning for image recognition. Its architecture consists of 19 layers, including 16 convolutional layers, 3 fully connected (FC) layers, and 5 max-pooling layers. The final layer is a softmax layer.

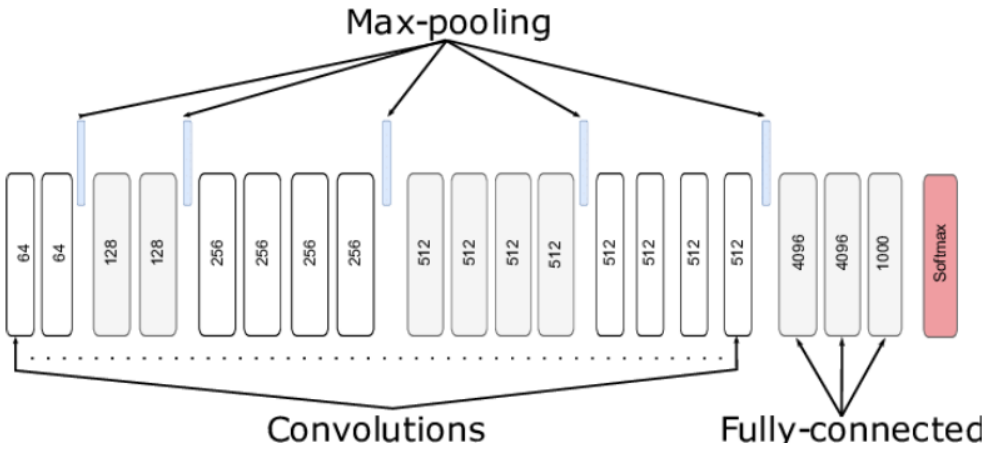


**Fig. 13**: VGG-19 architecture

## 5.5 DenseNet

A DenseNet is a type of convolutional neural network that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other. To preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers.

In a DenseNet architecture, each layer is connected to every other layer, hence the name Densely Connected Convolutional Network. For L layers, there are L(L+1)/2 direct connections. For each layer, the feature maps of all the preceding layers are used as inputs, and its own feature maps are used as input for each subsequent layer.

This is really it, as simple as this may sound, DenseNets essentially connect every layer to every other layer. This is the main idea that is extremely powerful. The input of a layer inside DenseNet is the concatenation of feature maps from previous layers.

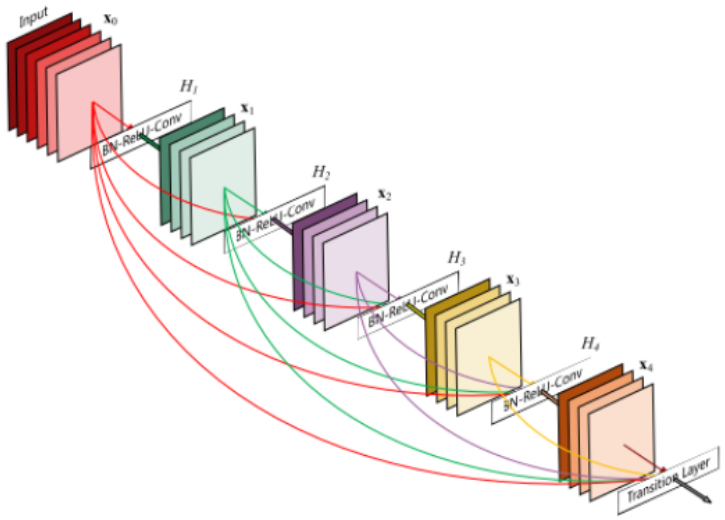


**Fig. 14**: DenseNet architecture

# 6 Experiment

## 6.1 Experiment Settings

We apply five models under five different preprocessing techniques to examine the impact of each method on models' ability to learn and classify diseases. Adam with a 3e-5 learning rate is used as an Optimizer. We use batch size of 8 and 10 epochs with BCEWithLogitsLoss as loss function. We then realize that with 10 epochs, SwinTran is still learning so we change its learning rate to 0.01 to witness its performance better.

## 6.2 Evaluation

As our task is performed in the medical domain, how to evaluate the performance of models and what metric to use are affected by many factors. However the most reasoning factor is the morality and the ethical issues in medic as one wrong decision could ruin somebody's life. So to ensure that we would satisfy those reasoning, we would use 4 metrics are: Precision, Recall, F1-score, Accuracy. Because we are working with an imbalanced dataset where all classes are equally important, using the macro average would be a good choice as it treats all classes equally.

### 6.2.1 Accuracy Score

The accuracy score is calculated as the ratio of correct predictions (both true positives and true negatives) to the total number of cases examined. Mathematically, it can be expressed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- True Positives (TP): The number of cases correctly identified as having the condition.
- True Negatives (TN): The number of cases correctly identified as not having the condition.
- False Positives (FP): The number of cases incorrectly identified as having the condition (also known as Type I error).
- False Negatives (FN): The number of cases incorrectly identified as not having the condition (also known as Type II error).

### 6.2.2 Precision Score

Precision measures the proportion of true positive predictions among all the positive predictions made by the model. It is calculated by dividing the number of true positives by the sum of true positives and false positives.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

### 6.2.3 Recall Score

Recall, also known as sensitivity, measures the proportion of true positive predictions among all the actual positive samples in the dataset. It is calculated by dividing the number of true positives by the sum of true positives and false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

### 6.2.4 F1-Score

F1-Score is a weighted average of Precision and Recall, where the weights are equal. It is used to balance the trade-off between precision and recall.

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4}$$

## 6.3 Results and Discussion

As we can see in Table 2, the SwinTransformer, Unetr and VGG19 are the highest performance models with the f1 score of 0.3657, 0.3609, 0.3617. Meanwhile, CoAtNet and DenseNet seem to not perform well on this task. These results proved the effectiveness of the Transformer architecture applied on the image classification tasks with 2 out of three Transformer architectural models having the high performance. Besides that, VGG19 surprised us by proving the effectiveness in extracting features from medical x-ray images and classified them. However, without the resources to continue training, we could not come to the conclusion of the performance of the models as the score was still increasing and the loss was still declining when we finished training at 10 epochs.

| Model | F1-macro | Precision | Recall | Accuracy |
|---|---|---|---|---|
| SwinTran | **0.3657** | 0.5332 | **0.3703** | 0.7872 |
| UNetr | 0.3609 | **0.5887** | 0.3196 | **0.8250** |
| CoAtNet | 0.2177 | 0.4025 | 0.2032 | 0.8049 |
| VGG19 | 0.3617 | 0.4642 | 0.3432 | 0.8058 |
| DenseNet | 0.1560 | 0.1811 | 0.1563 | 0.7914 |

**Table 2**: Overall result of the models without preprocessing

## 6.4 Analyzing the effectiveness of image preprocessing

| Methods | SwinTran | Unetr | CoAtNet | VGG19 | DenseNet |
|---|---|---|---|---|---|
| No preprocessing | 0.3657 | 0.3609 | 0.2177 | 0.3617 | 0.1560 |
| Exposure + Hist | 0.3783 | 0.3428 | 0.2175 | **0.3672** | **0.1582** |
| Laplace Gaussian | **0.3941** | **0.3625** | **0.2384** | 0.2164 | 0.1485 |
| Sobel | 0.3611 | 0.3436 | 0.2244 | 0.2550 | 0.1386 |
| Canny | 0.3576 | 0.3177 | 0.1494 | 0.2553 | 0.1397 |
| Thresholding - noisy | 0.1476 | 0.1043 | 0.1494 | 0.1494 | 0.1043 |
| Thresholding - less noisy | 0.1494 | 0.1043 | 0.1472 | 0.1494 | 0.1043 |

**Table 3**: Overall result of the models with five preprocessing methods

According to Table 3, Without applying any preprocessing technique, SwinTran and VGG19 perform similarly well with F1-macro scores of 0.3657 and 0.3617, respectively, while DenseNet has the lowest performance at 0.1560. Exposure + Histogram method helps all models except UNetr enhance their results and become the best method for VGG19 and DenseNet. On the other hand, Laplace Gaussian method significantly lowers the performance of these two models but it is the most suitable preprocessing technique for SwinTran, UNetr and CoAtNet. In contrast, Sobel methods lead to mixed results, with CoAtNet achieving slightly better scores (0.2244). Both thresholding methods make all five models unable to learn anything and drop their performance to the bottom with the result around 0.1043 to 0.1494.

To have a clearer understanding on what classes and labels the preprocessing methods affect, we displayed the scores on each label of the two highest performance models (SwinTran, VGG19) in table 4 . Observing from table 4, there were improvements in classification performance of 8 classes (Cardiomegaly, Consolidation, Infiltration, Lung Opacity, Nodule/Mass, Pleural effusion, Pleural thickening, Pulmonary fibrosis) when applying Laplace-Gaussian transform + Swin Transformer compared to the original Swin Transformer. However the Laplace-Gaussian also witnessed the reduction in performance on 3 classes (Atelectasis, ILD, Other lesion). We could explain that the preprocessing method is not an ideal way for all classes of diseases as the disease signals are located differently on the human bodies and using preprocessing

| Disease | SwinTran | | VGG19 | |
|---|---|---|---|---|
| | Original | Laplace-Gaussian | Original | Exposure + Histogram |
| Aortic enlargement | 0.85 | 0.85 | 0.85 | 0.84 |
| Atelectasis | 0.25 | 0.23 | 0.00 | **0.10** |
| Calcification | 0.03 | 0.03 | 0.07 | 0.03 |
| Cardiomegaly | 0.71 | **0.74** | 0.78 | 0.75 |
| Consolidation | 0.06 | **0.11** | 0.25 | 0.25 |
| Interstitial Lung Disease (ILD) | 0.43 | 0.34 | 0.09 | **0.16** |
| Infiltration | 0.03 | **0.30** | 0.29 | 0.25 |
| Lung Opacity | 0.52 | **0.54** | 0.46 | 0.40 |
| Nodule/Mass | 0.12 | **0.15** | 0.15 | **0.24** |
| Other lesion | 0.45 | 0.43 | 0.35 | 0.35 |
| Pleural effusion | 0.54 | **0.60** | 0.54 | 0.54 |
| Pleural thickening | 0.55 | **0.60** | 0.64 | **0.67** |
| Pneumothorax | 0.00 | 0.00 | 0.00 | 0.00 |
| Pulmonary fibrosis | 0.58 | **0.59** | 0.57 | 0.56 |

**Table 4**: F1-macro score of each label of SwinTran and VGG19 with their best preprocessing technique

methods might affect the signals of some specific diseases. Continuing with the result on VGG19 and VGG19 + Exposure, histogram, the performance of exposure and histogram technique seems to not be as good as the laplace gaussian one when the number of reduction in classification performance is much more than laplace's. There was reduction in 6 classes include (Aortic enlargement, Calcification, Cardiomegaly, Infiltration, Lung Opacity, Pulmonary fibrosis) while the amount showing positive performance are only 4 (Atelectasis, ILD, Nodule/Mass, Pleural thickening). Notably, neither of the preprocessing techniques can help two models detect the Pneumothorax label. The results showed an important problem when applying preprocessing methods is the suitability of classes, right methods and right classes would help in improving the performance. In conclusion, the experiments came up with two main ideas including that not all diseases need to be preprocessed as the preprocess might reduce the signals in images and not all the preprocessing techniques can be applied on a specific class of diseases.

After realizing the importance of the suitable preprocessing methods for disease classification, we analyzed deeper into the detailed performance of each method. We decided to choose the SwinTran model due to the highest effectiveness in this task. We calculated the f1-score on each class and presented the results in table 5. The method using noise thresholding and less noisy thresholding did not perform well on this task and caused the declining in its own performance so we would not include those two into the table. Firstly, we can immediately observe that all three preprocess methods increased the f1-score on at least three classes and lowered the f1-score of the remaining as well. However, according to class Atelectasis, we witnessed the worse performance on the most effective method LapGau and Sobel compared to the better result on using Histogram. It means that the Hist method is more suitable for this class rather than the other two; the same phenomenon appeared on Calcification, Module/Mass and Pulmonary fibrosis when Hist outperformed LapGau. Secondly,

| Disease | Original | Laplace-Gaussian | Exposure + Histogram | Sobel |
|---|---|---|---|---|
| Aortic enlargement | 0.85 | 0.85 | 0.85 | 0.83 |
| Atelectasis | 0.25 | 0.23 | **0.31** | 0.22 |
| Calcification | 0.03 | 0.03 | **0.07** | 0.00 |
| Cardiomegaly | 0.71 | **0.74** | **0.75** | 0.69 |
| Consolidation | 0.06 | **0.11** | 0.00 | 0.00 |
| Interstitial Lung Disease (ILD) | 0.43 | 0.34 | 0.37 | 0.33 |
| Infiltration | 0.03 | **0.30** | **0.22** | **0.24** |
| Lung Opacity | 0.52 | **0.54** | 0.51 | 0.51 |
| Nodule/Mass | 0.12 | **0.15** | **0.19** | 0.03 |
| Other lesion | 0.45 | 0.43 | 0.44 | 0.41 |
| Pleural effusion | 0.54 | **0.60** | 0.49 | 0.44 |
| Pleural thickening | 0.55 | **0.60** | 0.48 | **0.59** |
| Pneumothorax | 0.00 | 0.00 | 0.00 | **0.15** |
| Pulmonary fibrosis | 0.58 | **0.59** | **0.63** | **0.62** |

**Table 5**: F1-macro score of each label of SwinTran under different preprocess methods

although the Sobel method did not perform well, it outperformed LapGau on enhancing Pulmonary fibrosis classification. This has increased our trust that each class of disease will be suitable with a preprocessing method. And finally Pneumothorax, a disease which neither LapGau nor Hist came up with a solution for classifying, Sobel surprisingly stood up with the ability to handle the issue despite the worse overall performance when compared to the other two.

# 7 Conclusion and Future Work

In this report, we conduct some experiments with medical image preprocessing and deep learning architecture methods selection. Swin Transformer achieves highest F1-macro score of 0.3657 on the test set for chest X-ray abnormalities detection task without using any preprocessing technique. This result can be accepted because of the variety of the labels from the dataset. We also compare the impact of each preprocessing method on the models and find out that Laplace Gaussian is suitable with SwinTran, Unetr and CoAtNet, while Exposure + Histogram is more compatible with VGG19 and DenseNet. Swin Transformer with Laplace Gaussian and VGG19 with Exposure + Histogram has the best performance with an F1-macro score of 0.3941 and 0.3672, respectively.

We would like to conduct follow-up research to improve performance by acquiring more diverse chest X-ray images to enhance the robustness of the model and investigating instance segmentation techniques to identify and segment individual abnormalities within chest X-ray images.

# References

[1] Dai, Z., Liu, H., Le, Q.V., Tan, M.: Coatnet: Marrying convolution and attention for all data sizes. Advances in neural information processing systems **34**, 3965–3977 (2021)

[2] Baltruschat, I.M., Nickisch, H., Grass, M., Knopp, T., Saalbach, A.: Comparison of deep learning approaches for multi-label chest x-ray classification. CoRR **abs/1803.02315** (2018) 1803.02315

[3] Huang, G., Liu, Z., Maaten, L., Weinberger, K.Q.: Densely Connected Convolutional Networks (2018)

[4] Huang, X., Fang, Y., Gu, M.: Classification of chest x-ray disease based on convolutional neural network. Journal of System Simulation **32**(6), 1188 (2020)

[5] Li, Y., Zhang, Z., Dai, C., Dong, Q., Badrigilan, S.: Accuracy of deep learning for automated detection of pneumonia using chest x-ray images: A systematic review and meta-analysis. Computers in Biology and Medicine **123**, 103898 (2020) https://doi.org/10.1016/j.compbiomed.2020.103898

[6] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[7] Petit, O., Thome, N., Rambour, C., Soler, L.: U-Net Transformer: Self and Cross Attention for Medical Image Segmentation (2021)

[8] Sharma, S., Guleria, K.: A systematic literature review on deep learning approaches for pneumonia detection using chest x-ray images. Multimedia Tools and Applications **83**(8), 24101–24151 (2024)

[9] Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (2015)

[10] Taslimi, S., Taslimi, S., Fathi, N., Salehi, M., Rohban, M.H.: SwinCheX: Multi-label classification on chest X-ray images with transformers (2022)

[11] Zouch, W., Sagga, D., Echtioui, A., Khemakhem, R., Ghorbel, M., Mhiri, C., Hamida, A.B.: Detection of covid-19 from ct and chest x-ray images using deep learning models. Annals of Biomedical Engineering **50**(7), 825–835 (2022)

[12] Kanopoulos, N., Vasanthavada, N., Baker, R.L.: Design of an image edge detection filter using the sobel operator. IEEE Journal of solid-state circuits **23**(2), 358–367 (1988)

[13] Nguyen, H.Q., Lam, K., Le, L.T., Pham, H.H., Tran, D.Q., Nguyen, D.B., Le, D.D., Pham, C.M., Tong, H.T.T., Dinh, D.H., Do, C.D., Doan, L.T., Nguyen, C.N., Nguyen, B.T., Nguyen, Q.V., Hoang, A.D., Phan, H.N., Nguyen, A.T., Ho, P.H., Ngo, D.T., Nguyen, N.T., Nguyen, N.T., Dao, M., Vu, V.: VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations (2020)